standing Workshop, Monterey, CA, November 1994.

[Roberge 1985] J. Roberge, “A Data Reduction Algorithm for Planar Curves,” *Computer Vision, Graphics, and Image Processing*, Vol. 29, pp. 168-195, 1985.

[Rom 1993] H. Rom, *Part Decomposition and Shape Description*, Ph. D. Thesis, University of Southern California, December 1993.

[Rom & Médioni 1993] H. Rom and G. Médioni, “Hierarchical decomposition and axial shape description,” In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 10, pp. 973-981, October 1993.

[Rom & Médioni 1994] H. Rom and G. Médioni, “Part Decomposition and Description of 3D Shapes,” In *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Slade 1991] S. Slade, “Case-Based Reasoning: A Research Paradigm,” In *AI Magazine*, pp. 42-55, Spring 1991.

[Winston 1975] P. H. Winston, “Learning Structural Descriptions from Examples,” in *The Psychology of Computer Vision*, P. H. Winston, Ed., New York:McGraw Hill, 1975, Chapter 5.

[Zerroug 1994] M. Zerroug, *Segmentation and Inference of 3-D Descriptions from an Intensity Image*, Ph. D. Thesis, University of Southern California, June 1994.

[Zerroug & Nevatia 1993a] M. Zerroug and R. Nevatia, “Scene Segmentation and Volumetric Descriptions of SHGCs from a Single Intensity Image,” In *Proceedings of the ARPA Image Understanding Workshop*, Washington, DC, April 1993.

[Zerroug & Nevatia 1993b] M. Zerroug and R. Nevatia, “Quasi-Invariant Properties and 3-D Shape Recovery of Non-Straight, Non-Constant Generalized Cylinders,” In *Proceedings of the ARPA Image Understanding Workshop*, Washington, DC, April 1993.

[Zerroug & Nevatia 1994a] M. Zerroug and R.Nevatia, “From an Intensity Image to 3-D Segmented Descriptions.” in *Proceedings of the IEEE International Conference on Pattern Recognition*, Jerusalem 1994.

[Zerroug & Nevatia 1994b] M. Zerroug and R.Nevatia, “Three-Dimensional Part-Based Descriptions from a Real Intensity Image,” in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

Detection," *Proceedings of the 1994 ARPA Image Understanding Workshop, Monterey,* California, November 1994.

[Chen 1994] Y. Chen, *Description of Complex Objects from Multiple Range Images*, Ph. D. Thesis, University of Southern California, August 1994.

[Chen & Medioni 1994a] Y. Chen and G. Medioni, "Shape Description of Complex Objects from Multiple Range Images", *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, Seattle WA., pp.153-158.

[Chen & Medioni 1994b] Y. Chen and G. Medioni, "Surface Description of Objects from Multiple Range Images," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Clemens & Jacobs 1991] D. T. Clemens and D. W. Jacobs, "Space and Time Bounds on Indexing 3-D Models from 2-D Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1991, pp. 1007-1017

[Draper *et al* 1989] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman, "The Schema System," *International Journal of Computer Vision, Vol. 2*, No. 3, 1989, pp. 209-250.

[Grimson 1990] W. Grimson, *Recognition by Computer-The Role of Geometric Constraints*, Cambridge, MA:MIT Press, 1990.

[Guy & Médioni 1994] G. Guy and G. Médioni, "Inferring Surfaces from Sparse 3-D Data," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Havaldar, *et al.* 1994] P.Havaldar, G.Médioni and F. Stein, "Extraction of Groups for Recognition," in *Proc. ECCV94*, Stockholm, May 1994.

[Huertas *et al.* 1990] A. Huertas, W. Cole, and R. Nevatia, "Detecting Runways in Complex Airport Scenes," *Computer Vision, Graphics, and Image Processing, Vol. 51*, No. 2, pp. 107–145, August 1990.

[Khokhar et al. 1993] A. Khokhar and H-J. Kim and V. Prasanna, "Scalable Geometric Hashing on MasPar MP-1," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, New York, NY, pp. 594-595, June 1993.

[Kim 1993] Y. C. Kim, *Integrated Motion Analysis System Guided by Feedback Information*, Ph. D. Thesis, University of Southern California, August 1993.

[Kim 1994] D. Kim, *Indoor Navigation with a Generic Map*, Ph. D. Thesis, University of Southern California, 1994.

[Kim & Nevatia 1993] D. Kim and R. Nevatia, "Indoor Navigation without a Specific Map," in *Proceedings Workshop on Intelligent Systems*, Pittsburgh, PA, 1993

[Kim & Nevatia 1994] D. Kim and R. Nevatia, "A Method for Recognition and Localization of Generic Objects for Indoor Navigation," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Kolodner 1992] J. Kolodner, "An Introduction to Case-Based Reasoning," in *Artificial Intelligence Review*, Vol. 6, pp. 3-34, 1992.

[Liao & Médioni 1994] C. Liao and G. Médioni, "Surface Approximation of Complex 3-D Objects," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Lin *et al.* 1994] C. Lin, A. Huertas and R. Nevatia, "Detection of Buildings Using Perceptual Grouping and Shadows", Proceedings of the *IEEE Computer Vision and Pattern Recognition Conference*, Seattle, June 94, pp. 62-69

[MacGregor & Burstein 1991] R. MacGregor and M. Burstein, "Using a Description Classifier to Enhance Knowledge Representation," *IEEE Expert, Vol 6,* No. 3, pp. 41-46, June 1991

[McKeown *et al.* 1985] D. M. McKeown, Jr., W. A. Harvey, and J. McDermott, "Rule Based Interpretation of Aerial Imagery," *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 7,* No. 5, September. 1985, pp. 570-585.

[Milhaud & Médioni 1994] N. Milhaud and G. Médioni, "Learning, Recognition and Navigation from a Sequence of Infrared Images," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Mundy *et al.* 1993] J. Mundy and The Gang of Ten, "The Image Understanding Environment: Overview," in *Proc. ARPA Image Understanding Workshop*, Washington, DC, April 1993, pp. 283-288.

[Nevatia & Babu 1980] R. Nevatia and K. Babu, "Linear Feature Extraction and Description," *Computer Graphics and Image Processing*, Vol. 13, pp. 257-269, 1980.

[Prasanna & Wang 1994] V. Prasanna and C.-L. Wang, "Image Feature Extraction on Connection Machine CM-5," in *Proc. ARPA Image Understanding Workshop*, Monterey, CA, November 1994.

[Prasanna et al. 1993] V. Prasanna, C.-L. Wang, and A. Khokhar, "Low Level Vision Processing on Connection Machine CM-5," in *Workshop on Computer Architectures for Machine Perception*, pp. 117-126, 1993.

[Price *et al.* 1994] K. Price, T. Russ and R. MacGregor, "Knowledge Representation for Computer Vision: The Veil Project," in *Proc. ARPA Image Under-*

knowledge representation technology can be applied to computer vision and to explore extensions to knowledge representation systems to directly aid computer vision research (especially in the area of spatial reasoning).

The VEIL project investigates the benefits available to vision applications obtainable via the introduction of declarative programming techniques, specifically, techniques available using advanced symbolic processing technology found in a modern knowledge representation system.

In typical vision applications today, a programmer invents specialized data structures and carefully crafts a suite of vision processing algorithms that exploit those data structures. The result is most often a highly specialized piece of code that cannot be reused for a different domain, or applied to applications other than the one originally intended. The Image Understanding Environment project [Mundy *et al.* 1993] addresses the issue of basic data structures and processing but does not deal with higher level representation issues that are addressed here.

This project aims to develop a technology whereby much of the work that goes into the development of specialized vision processing modules results in software that can be shared or reused by multiple applications. Knowledge representation techniques have been a part of computer vision from the beginning [Winston 1975, McKeown *et al.* 1985, Draper *et al* 1989]. The difference in this project includes the power of the knowledge representation system being used and the use of relatively mature computer vision programs and techniques as the basis for incorporating knowledge representation technology. This work is reported more fully in [Price *et al.* 1994].

8 Parallel Implementation of Algorithms

We have recently initiated a project on parallel implementation of some of our IU algorithms on a Thinking Machines CM-5 computer. Our first task is parallel implementation of the monocular building detection system described in Section 6. The first step of this system, as well as many of our other IU systems, is linear feature extraction. Even though linear feature extraction is considered a low level task, we know of no parallel system implementation for this task. This has been the focus of our initial effort.

Linear feature extraction consists of several stages: edge detection, edge linking and line approximation. Edge detection is the process that has been typically studied for parallel implementations. The other steps, those of linking and approximation, however pose some challenges as the data structures change from being “iconic” to symbolic (lists). We have successfully implemented a version of the USC line finder [Nevatia & Babu 1980] on a CM-5.

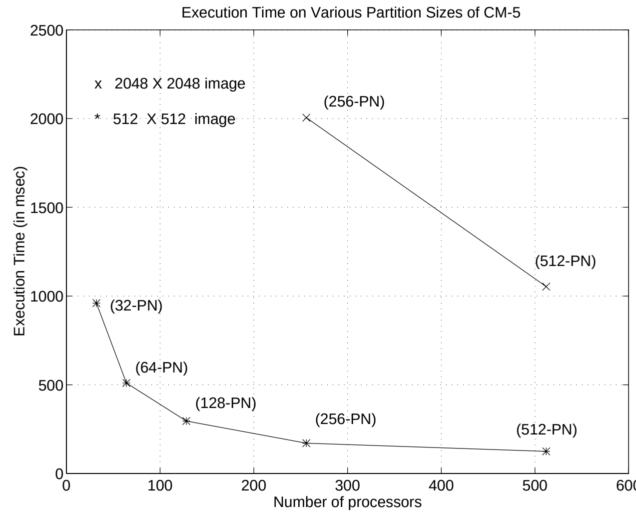


Figure 15 The execution time of the line finder for various sizes of image on various partition sizes of CM-5.

In Figure 13, execution times of the low level system on various image sizes are plotted against different partition sizes. The largest image we have processed is of size 2048X2048. The total execution time for processing such an image (including contour detection and linear approximation) is 1.053 seconds. The same image processed by a Sun Sparc 400 station operating at 32 MHz using a optimized code written in C takes more than 8 minutes. The times reported are the total CPU time measured by the CM-5 timer which has a resolution of 1 microsecond. The resolution of the clock on Sparc 400 is 16.667 milliseconds. All the reported times do not consider I/O time. This effort is described in more detail in [Prasanna & Wang 1994].

9 ACKNOWLEDGMENTS

This paper represents the work of many different current and former students, visitors, staff members and associated faculty. M. Zerroug worked on object descriptions from intensity images. Y. Chen and C. Liao worked on analysis of range images. H. Rom worked on volumetric descriptions of 3-D objects. G. Guy worked on inference of surfaces using perceptual grouping. P. Havaldar worked on perceptual grouping for object recognition. A. Francois worked on case-based reasoning for recognition. D. Kim worked on indoor navigation. N. Milhaud worked on navigation from infrared images. A. Huertas, C. Lin and M. Bejanin worked on the aerial image analysis. V. Prasanna and C. Wang worked on the analysis of parallel algorithms.

REFERENCES

[Bejanin, *et al.* 1994] M. Bejanin, A. Huertas, G. Medioni and R. Nevatia, “Model Validation for Change

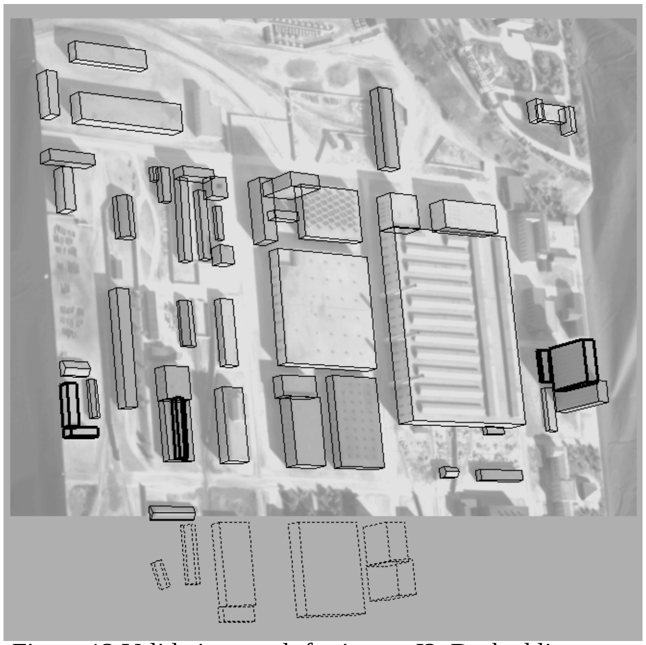


Figure 12 Validation result for image J2. Dashed lines (objects outside the field of view) and bold lines denote the few buildings NOT validated.

Our system is designed to validate buildings, which comprise the bulk of a site model. An example of this system is shown in Figure 12. This capability has been tested on the available model board imagery. Details are given in a paper [Bejanin *et al.* 1994] in these proceedings.

6.2 Building Detection

We are currently studying detection of objects in the image that are not in the site model. This step requires site modeling capabilities such as those provided by our monocular building detection system [Lin *et al.* 1994]. This system is restricted to rectangular shapes and uses shadows. Rectangular shapes are found by a perceptual grouping procedure using linear features extracted from the scene. These are verified using the shadow information and heights are approximated using the shadows. The current system handles images containing multiple buildings. Performance is excellent for nadir views. Figure 13 show a a sample of the results. The scene contains about 41 visible buildings, 34 (83%) of which are detected and verified by shadows. The remaining buildings, or portions of them, are actually hypothesized but lack sufficient shadow evidence. There are two false detections: one corresponds to a rectangular fenced area having supporting shadow evidence, and the other to a rectangular area between two parallel and aligned buildings. Dark buildings continue to be a source of difficulty as some of their boundaries may lack sufficient contrast with the adjacent shadow.

Oblique views are also handled, but further development is necessary for similar quality results for a wide variety of viewing conditions. Figure 14 shows the buildings detected from a portion of an image of an oblique view of the modelboard scene. The visible walls of the buildings are explicitly detected and help, together with the shadow evidence, verify and derive a 3-D description of the buildings. The detected 3-D structures are shown on the figure in black.

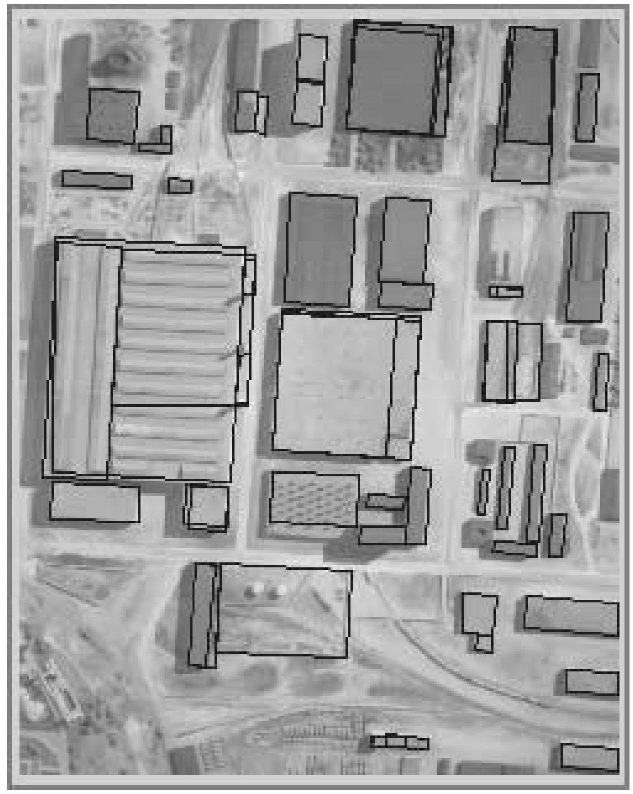


Figure 13 Automatically detected buildings from RADIUS modelboard image J3.

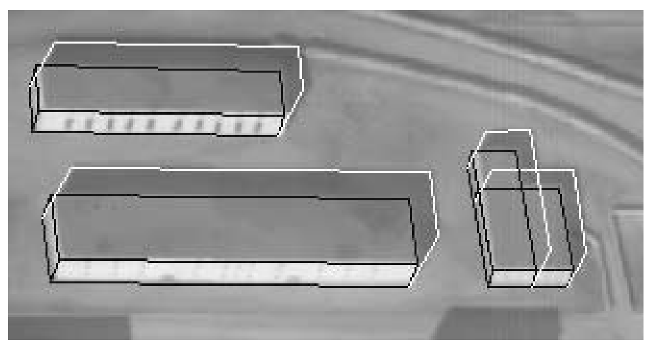


Figure 14 Building detection for oblique views. The detected 3-D buildings are marked in black with the shadows marked in white.

7 Knowledge Representation in Computer Vision

The VEIL (Vision Environment Integrated with Loom) project is a collaborative effort with ISI's Loom project (Knowledge Representation) [MacGregor & Burstein 1991]. It is an experiment both to see how traditional

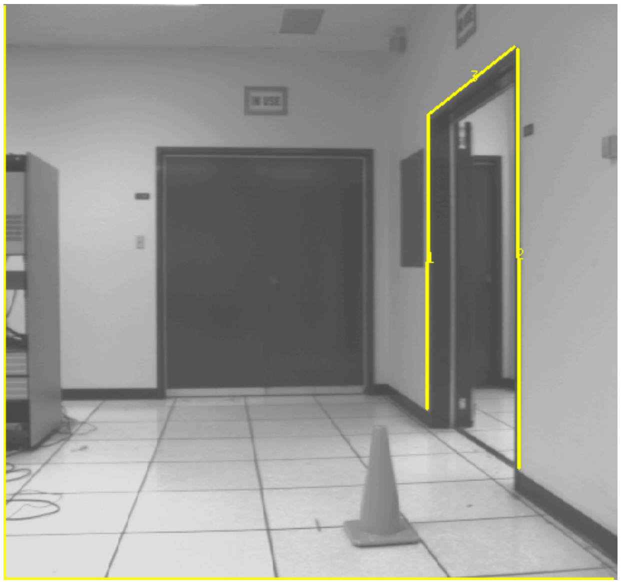


Figure 9 Indoor scene used for navigation with detected door outlined.

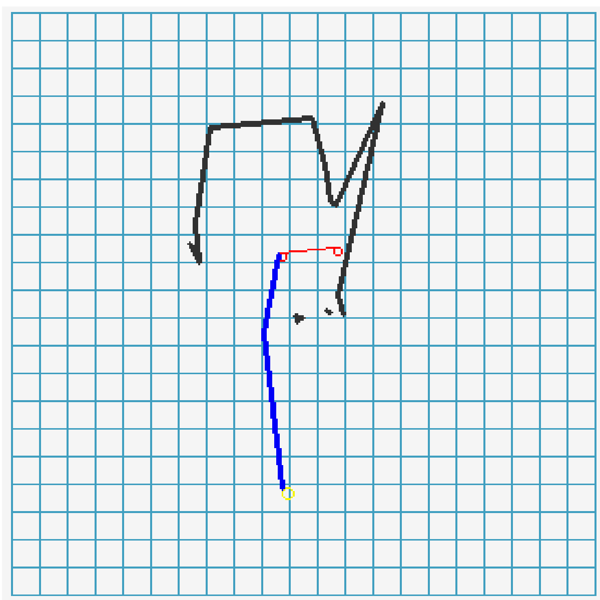


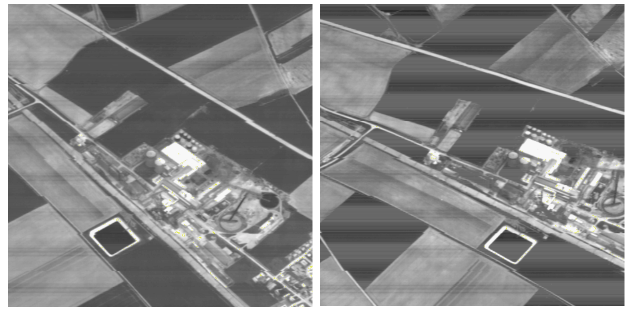
Figure 10 Map constructed from the scene used for robot path planning.

feasible (on special hardware). In this paper, we describe the software version of such an automated system This process in illustrated in Figure 11, which shows the 3D reconstruction from a sequence of images and the predicted appearance of a structure. This work is described in more detail in [Milhaud & Medioni 1994].

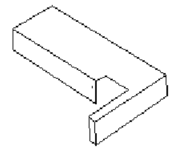
6 Aerial Image Analysis (RADIUS)

Our research effort in support of the RADIUS program consists of two projects:

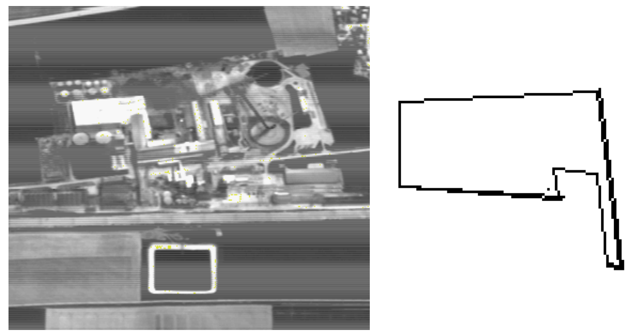
1) Research in Model-based change detection and site model updating.



a. Two images from the sequence



b. Reconstructed 3-D building



c. Image from unknown viewpoint and model predicted appearance

Figure 11 Illustration of recognition and navigation from Infrared Images.

2) Automatic building detection and description from monocular images.

6.1 Model-Based Change Detection

The objective of our first effort is to detect significant changes in a site from the last look and an updating of the site model to reflect these changes. Change detection is one of the most critical function in photo-intelligence analysis.

We view the task of change detection as consisting of three steps: detection, description and functional inference. Our approach to detection is by matching site model to descriptions generated from the new image at several levels. Missing model features and new image features indicate change. Descriptions of change will be required to distinguish significant changes from incidental ones and to provide input for functional inference.

We have made significant progress in the first part of our task, that of matching site models to images. Our system matches site models to image features for the purpose of “model validation”, that is verifying whether the objects in the site model are still present in an image.

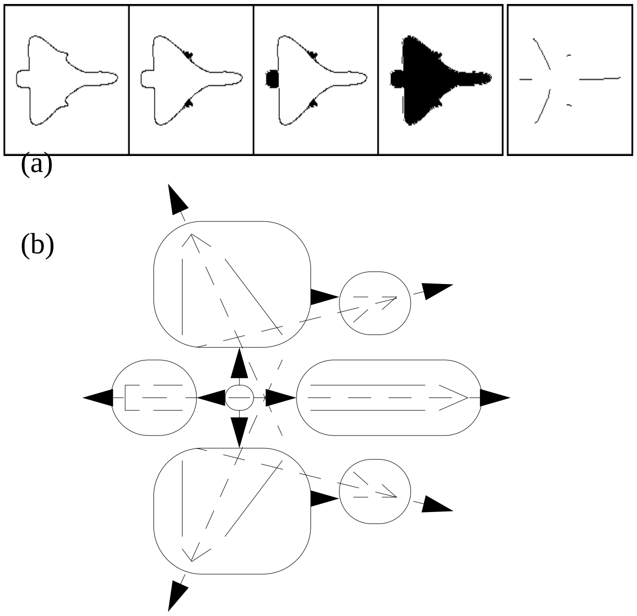


Figure 8 “F106” shape description obtained from real data. (a) Decomposition and (b) internal representation

We have developed the recognition engine by borrowing ideas from the Case-Based Reasoning (CBR) paradigm [Slade 1991, Kolodner 1992], which embeds all the characteristics required by our goals, such as the ability to process noisy and incomplete data, and provides an interesting framework for further intelligent processing (e.g. justified interpretation, automatic learning). However, we currently focus only on the low-level aspects of the reasoning: the description-storage-retrieval process.

Actual shape descriptions are stored in a data-base, from which they must be efficiently retrieved when a new shape is proposed for recognition. This representation is illustrated in Figure 8. We define a hierarchical indexing system based on the structure of the descriptions and the local description of parts (subset of the parameters describing the parts). This mechanism allows for dynamic updating of the data-base with a minimum computing cost to update the index.

The first process when a new shape is submitted for recognition is the retrieval of the closest shapes stored in the data-base. A partial match, based on the connection structure and the aggregation of dissimilarities between parts, is computed incrementally level by level between the new shape and the possible candidates. The combination of the incremental process with the hierarchical indexing, makes the number of shapes processed at each step decrease rapidly, therefore dramatically reducing the average complexity of the retrieval. The selected retrieved shape(s) are used to give a classification for the new shape.

We describe our implementation for 2D shapes recognition using the description process proposed in [Rom & Medioni 1993]. Among the perspectives for our system are its adaptation to 3D shapes based on the work described in [Zerroug 1994], which only requires the redefinition of the descriptive model without any modification of the algorithms, and the implementation of the high-level intelligent steps of the CBR recognition process.

5 Indoor Navigation and Dynamic Scene Analysis

We have been developing a vision system for indoor robot navigation. Problems of indoor navigation are quite different from those of outdoor navigation. The surface on which a robot must move is usually flat and smooth. However, the environment can be much more cluttered and variable. We believe that for such a task, it is not, in general, feasible to use a detailed map of the environment. Instead, we can only provide general descriptions of the likely object, such as characteristics of a room and pieces of furniture in it. Besides this requirement, we have also chosen to use a trinocular camera system as our sensor, in preference to an active range sensor. Range sensors are still quite expensive and do not work effectively for all kinds of surfaces encountered in an indoor environment.

In earlier work, we described the ability of our system to navigate in corridors in presence of obstacles [Kim & Nevatia 1993]. Here, the robot is not given a detailed map of the corridors, but simply a description of the corridors defined by vertical walls and a horizontal floor.

In recent work, we have been focusing on navigation in our laboratory. It contains a variety of objects such as a computer chassis, desks, walls and doors. Our objective is to have the robot reach goals defined by commands such as “go past the desk on your right and go through the first open door”. To do this, the robot must be able to recognize generic objects such as desks and doors. We have developed a system to such recognition and it is described in more detail in another paper in these proceedings [Kim & Nevatia 1994]. An example of a typical scene is shown in Figure 9. Note that the detected door is shown in white outline. Figure 10 shows a map we construct from the scene which is used for motion planning for the robot.

In a second project, we are addressing the problem in which an autonomous system equipped with a single infrared camera “learns” a designated rigid 3-D scene so that, at another time, it can recognize it and guide itself relative to the reconstructed scene, starting from an approximately known viewpoint, to reach a given destination. This scenario is relevant to several domains, in particular military missions and robotic navigation. A goal of our system is that a real-time implementation be actly known rigid object in a scene. The tools used to achieve this task are geometric constraints, and a lucid treatment of this class of approaches can be found in Grimson's book [Grimson 1990]. The presence of a model is inferred by the verification that such a model could indeed produce some of the observed data under an appropriate geometric transform. In this approach, low level primitives such as edgels or their approximations, as produced by state of the art edge finders can be used and an exact geometric model is required. Such an approach is therefore very appropriate when evolving in a controlled environment, such as a factory, where the number of possible objects is small and their geometry is precisely known.

However, this approach cannot be extended to more general scenarios because objects may be very similar while being geometrically different. Consider for instance two different airplanes which have similar features but different geometries. In other words, generic recognition should not make use of methods based purely on the exact geometric structure of the object. It is clear that the only way to solve this difficult problem is to reason about parts and their arrangements. To achieve this difficult goal, we therefore propose to:

- generate rich, stable descriptions from images, and to use perceptual grouping laws to achieve this task;
- develop a recognition engine with the ability to process noisy and incomplete data. We are exploring the use of Case Based Reasoning to this effect.

We now describe in more detail these two aspects.

4.1 Perceptual Grouping for Generic Object Recognition

Generic object recognition deals with recognizing certain classes of objects. It requires models to be defined in terms of high level parts and their arrangement, instead of their exact geometry.

We propose to use perceptual grouping laws to extract groups from initial primitives, organize them into sets which have similar "perceptual" content and use the sets for recognition. We use multiple views of an object as our models. Elements in the edge image are grouped based on proximity, similarity, closure, symmetry, and continuation. First, we generate features with multiple representations to overcome the unreliability of local algorithms during preprocessing, and to handle noise and capture different levels of detail. We then construct a hierarchy of groups based on various grouping criteria.

The groups, as extracted from real images, not only contain perceptually salient groups but also accidental groups which do not yield any natural descriptions and increase the complexity of the recognition process. We therefore propose a way to select "good" groups and if possible, discard the "bad" groups.

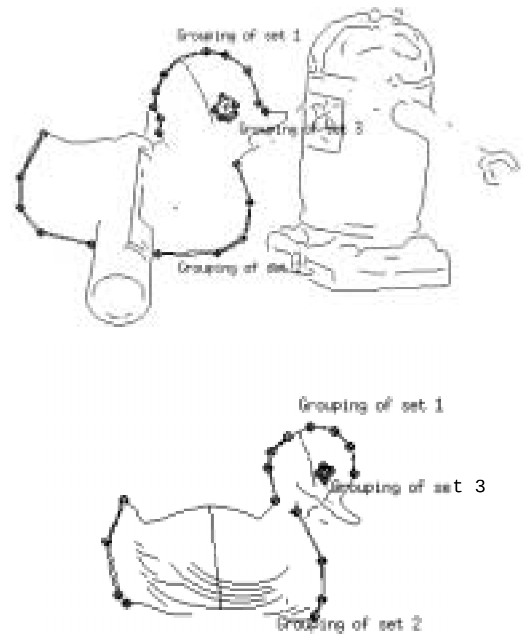


Figure 7 Example of Generic Recognition showing detected similar features. The Duck in scene (top) is different from that in model (bottom).

For recognition, we follow a graph-based matching approach. We use a set of non-registered views of a 3-D object as models. We use sets of perceptually similar groups to model the object and perform recognition. As such, every valid group that the object gives rise to has *multiple representations*. While most other systems use spatial correspondences to verify matching hypotheses, we use high level features and their *topological* relationships for the recognition process. These topological relationships can be represented by a graph. The graph vertices represent sets of features. The edges represent topological attributes between the features, such as inclusion and adjacency. We use the paths in this graph as basic token for the process of structural indexing. Groups of consistent hypotheses represent detected model instances in a scene.

4.2 Hierarchial Indexing

In this work, we address the problem of human-like generic shape recognition, which we define as a classification rather than an identification problem.

Our description model assumes that a shape is already segmented into parts. The local geometrical description of parts is qualitative, and the connection parameters are normalized to be scale-independent. The described parts are organized into a connectivity acyclic oriented graph (the edges are oriented from the "biggest" to the "smallest" part).

Figure 4 Shaded range image of a teapot for volumetric descriptions.

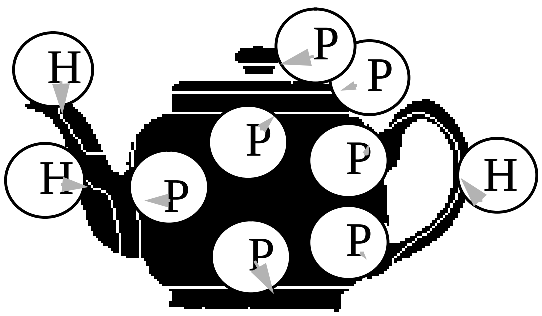


Figure 5 The parabolic curves of the teapot and their volumetric classification. P = “Part-like” and H = “Hump-like.”

3.3 Volumetric Descriptions of 3-D Objects

Surface descriptions from range data do not capture aspects of shape such as volumetric parts, which should help in recognition. We address the problem of obtaining natural (intuitive) descriptions of 3-D shapes of three dimensional compound objects, where the parts are connected smoothly. The input we consider is either complete 3-D data or range data from a single view. We suggest a *volumetric* graph representation of the object, where the nodes represent individual parts and the edges represent connectivity information. We suggest the use of properties of the parabolic curves for performing the part decomposition. We currently consider parts with tubular structure with a straight or curved axis. The graph description presents a structural description of the shape in terms of parts and their arrangement. We are also interested in the internal description of the parts. We suggest the use of properties of the parabolic curves for recovering natural descriptions of large classes of Generalized Cylinders in terms of their cross sections and axes. The more complete description is found in [Rom & Medioni 1994].

3.4 Surface Information from Sparse 3-D Data

We are addressing the problem of obtaining surface information from a sparse sets of 3-D data. This input can be in the form of points, or points with an associated normal, allowing for both position and direction to be corrupted by noise.

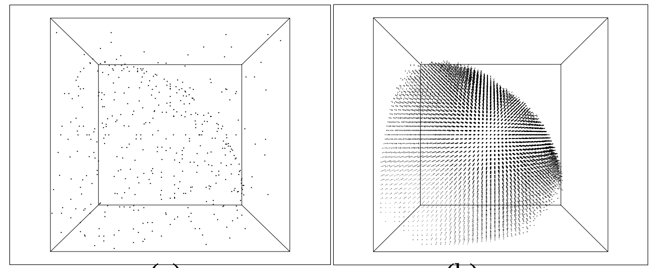


Figure 6 Surface description from sparse point data. (a) Input sparse point image. (b) Final saliency map (thresholded) showing the surface.

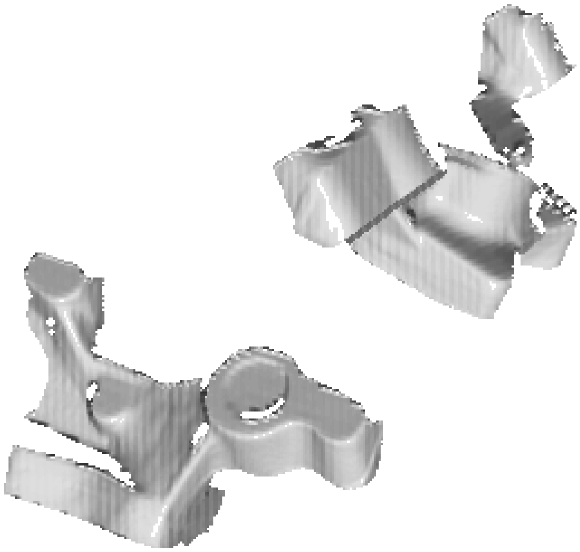
This is the typical input obtained from matching sparse features in stereo or motion, assuming that the observed scene is rigid. Most approaches treat the problem as an interpolation problem, solved by fitting a surface such as a membrane or thin plate which minimizes some functional. We argue that these physical constraints are not sufficient and propose to impose perceptual constraints such as good continuation and “co-surfacity.” These perceptual constraints allow us not only to infer surfaces, but also detect surface discontinuities at the same time.

We are able to handle scenes of any genus, any number of discontinuities, and of any number of objects, without a priori knowledge or special considerations. The result is in the form of three dense saliency maps for surfaces, intersections between surfaces, and 3-D junctions. These saliency maps can then be used to guide a ‘marching’ process to generate a CAD model of surfaces, space curves, and 3-D junctions.

The saliency maps are generated by convolving each feature with a pre-defined 3-D mask (later referred to as the Extension Field) resulting in a dense saliency map of the space. Such a map holds high values for locations in space which are strong candidates for surfaces. Also, at each such location, a normal to the predicted surface in available. Space curves and 3-D junctions can be easily derived from the same map, as we show later on. In Figure 6 we show an example where the input consists of a cloud of non-oriented points, with a considerable amount of noise. An octant of a sphere is embedded in noise. The sphere has ~200 data points, and ~100 noise points. The resulting saliency map (Figure 6(b)) delivers a dense map of normals in the computed location of the surface, while eliminating all traces of the noise points. The more complete description of this work is in [Guy & Medioni 1994].

4 Object Recognition

Most object recognition systems today address the problem of finding the location and orientation of an ex-



a. Two of the original range images.

b. View of the reconstructed model.

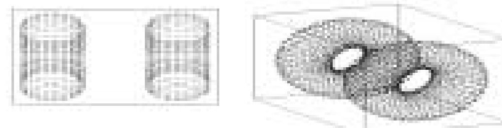
Figure 2 Reconstruction of model of an automobile part using the inflating balloon model from several range views.

on a carefully selected initial state or encountering local minimum problem. It also allows us to adapt the mesh surface to changes in local surface shapes and to handle holes present in the input data through adjusting certain system parameters adaptively and locally.An example of the reconstructed model of an automobile part is shown in Figure 2. This work is described in [Chen & Medioni 1994a, Chen & Medioni 1994b].

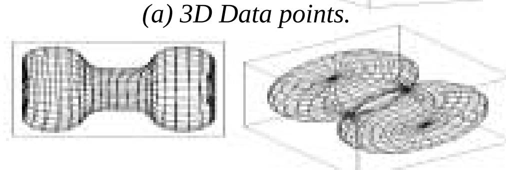
3.2 Surface Description and Segmentation of Complex Objects

Our goal is to generate a surface description of complex objects with parts and holes. We start by fitting a surface, assuming the object is of Genus 0, then analyze the result to further segment the description.

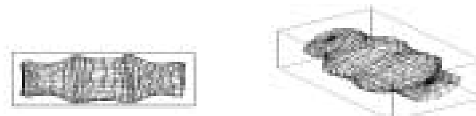
In the first part of our algorithm, the system provides an initial estimated surface which is subject to internal forces (describing implicit continuity properties



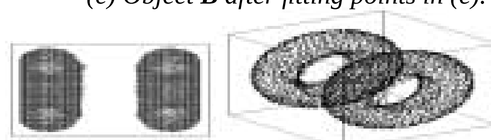
(a) 3D Data points.



*(b) Object **A** after the first fit.*



*(c) Object **B** after fitting points in (e).*



*(d) Result of the Boolean operation **A\B**.*

Figure 3 Surface description of complex objects, two views for each set of data are shown for an example with two tori.

such as smoothness) and external forces which attract it toward the data points. The problem is cast in terms of energy minimization. We solve this non-convex optimization problem by using the well known Powell algorithm which guarantees convergence and does not require gradient information. The variables are the positions of the control points. The number of control points processed by Powell at one time is controlled. This methodology leads to a reasonable complexity, robustness, and good numerical stability. We keep the time and space complexities in check through a coarse to fine approach and a partitioning scheme. We handle closed surfaces by decomposing an object into two caps and an open cylinder, smoothly connected. The process is controlled by two parameters only, which are constant for all our experiments.

Most deformable models are inadequate for objects more complex than Genus 0, with deep cavities. Furthermore, they always assume there is only one underlying object for the collected data, which means the segmentation, which is difficult, has been done ahead of time. In the second part, we propose an approach that can apply simultaneously more than one curve or surface to approximate multiple objects. Using the residual data points, the bad parts of the fitting curve, and appropriate Boolean operations, our approach is able to handle objects more complicated than Genus 0 or with deep cavities, and can perform segmentation if there is more than one underlying object. An example is shown in Figure 3. More details are found in [Liao & Medioni 1994].

This method is an extension of our previous work [Zerroug & Nevatia 1993a, Zerroug & Nevatia 1993b] which addressed single-part objects. In recent work, additional capabilities have been included in order to handle multi-part objects. They include the analysis of joint relationships and their effects in refining the segmentation and in recovering 3-D shape.

An example is given in Figure 1. Figure 1(a) shows the intensity image of a tea pot with a textured background. Figure 1(b) shows the edges detected in it. Note that the object boundaries are fragmented and that many extraneous boundaries due to markings and noise are present. Our objective is to detect the tea pot from the background and describe it in terms of its parts: the body of the pot, the lid, the handle and the spout. Our system, described more fully in [Zerroug & Nevatia 1994] does this; the results are shown in Figure 1(c). The figure shows the graphical representation of the segmented compound object in terms of its parts (nodes) and the joints (arcs). The nodes are labeled by the type of generalized cylinders that represent them best, including the GC elements (cross-section, axis and sweep) which are either 3-D if the 3-D recovery is possible or otherwise projective (in the orthographic sense). The arcs are labeled partially by the types of joints they represent. For this object, four joints have been detected.

Our system has been tested on a limited set of images only at this time. However, we believe that the results are promising and that the generated descriptions can be readily used for object recognition.

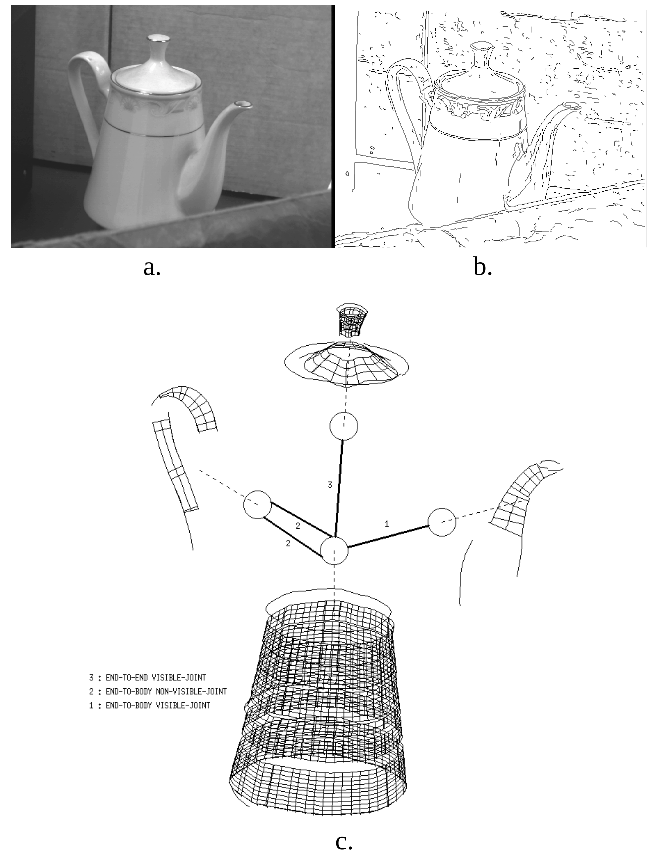


Figure 1 Results of object descriptions from intensity data. (a) sample intensity image; (b) its extracted edges; (c) the resulting compound object representation.

3 Analysis of Range Images

The goals of our effort in range image understanding are to generate rich descriptions from sensed 3-D data. These descriptions should be segmented and capture both the volumetric and surface information related to objects. One of the applications of the described research is the automatic generation of 3-D models from multiple range images. We describe below in more detail three specific aspects of our research, first the integration of multiple range images into a surface representation of the object, then a framework to handle complex objects with holes or multiple components, and an approach to generate volumetric part descriptions from a surface representation of an object.

In a related, but different, project we study the inference of surface shape from sparse three-dimensional data using perceptual organization

3.1 Surface Fitting with an Inflating Balloon Model

The goal here is to construct models of existing objects from their range images. This is important since we may not have access to models of all objects that a vision system needs to deal with and it is necessary to create such models from real objects. The model construction process can also be used in reverse engineering and CAD/CAM systems.

Our target is to develop a method that can be used to describe complex (i.e. non-star-shaped) objects, since most simple object cases are trivial. Existing shape description methods (such at deformable models) suffer from local minimum problem and hence usually depend upon a carefully selected initial guess to come up a good description. Our approach is based on a dynamic balloon model represented by using a triangulated mesh. The shape description process simulates blowing into a balloon inside the shell of the object. The vertices in the dynamic mesh are linked to their neighboring vertices through springs to simulate the surface tension and to keep the shell smooth. Unlike other dynamic models proposed by previous researchers, our balloon model is purely driven by an inflation force applied towards the object surface from inside of the object, until the mesh elements reach the object surface. The system includes an adaptive local triangle mesh subdivision scheme that results in an evenly distributed mesh. Since our approach is not based on global minimization, it can handle complex, non-star-shaped objects without relying

- the inference of surface shape from sparse three-dimensional data using perceptual organization (in Section 3.4).

Object Recognition

Most object recognition systems today address the problem of finding the location and orientation of an exactly known rigid object in a scene.However, these approaches cannot be extended to more general scenarios because objects may be very similar while being geometrically different. Consider for instance two different airplanes which have similar features but different geometries. In other words, generic recognition should not make use of methods based purely on the exact geometric structure of the object. It is clear that the only way to solve this difficult problem is to reason about parts and their arrangements.To achieve this difficult goal, we conduct two efforts to:

- generate rich, stable descriptions from images, and to use perceptual grouping laws to achieve this task (see Section 4.1)
- develop a recognition engine with the ability to process noisy and incomplete data. We are exploring the use of Case Based Reasoning to this effect (see Section 4.2).

Indoor Navigation and Dynamic Scene Analysis

We have been developing a vision system for indoor robot navigation. This system is based on a Denning mobile robot with a trinocular vision system. Our objective is to use generic descriptions of the path (go past the desk on your right and go through the first open door) rather than a detailed specific map. Currently our robot is able to navigate in laboratory environments avoiding obstacles and using objects such as doors and desks for landmarks.

We have also studied the problem of navigation in an environment which we model as we go. Specifically, we consider the problem of building a model of a scene, so that a similar sensor could, at a later time, orient itself with respect to this representation. See Section 5 for more information on both of these efforts.

Aerial Image Analysis (RADIUS)

Our work in aerial image analysis is currently supported under the RADIUS program. This research includes model-based change detection applied to buildings in the scene and monocular detection and analysis of building. These are discussed in Section 6.

Knowledge Representation in Computer Vision

This work focuses on integrating advanced knowledge representation technology (as represented by Loom) with image understanding technology to develop advanced tools for the generation of vision systems. It is aimed at enhancing current knowledge representation technology to support computer vision tasks in the realm of spatial reasoning and bringing modern knowledge representation technology to computer vision systems through the use of higher level representations. The result will improve shareability and reusability of code for computer vision systems. (See Section 7.)

Parallel Implementation of Algorithms

We are exploring the implementation of standard computer vision algorithms on parallel machines [Prasanna & Wang 1994, Prasanna *et al.* 1993]. Our recent experiments on implementing a our line finder system on a Thinking Machines CM-5 are discussed in Section 8.

PH. D. Graduates

A major product of our research is graduates. In this period we have had Y. C. Kim [Kim 1993], H. Rom [Rom 1993], Y. Chen [Chen 1994], M. Zerroug [Zerroug 1994], and D. Kim [Kim 1994].

Further Information

We make descriptions of most of our research efforts available on-line through the WWW file: “http://iris.usc.edu/USC-Computer-Vision.html.” This file has current summaries of our research, example results, some recent papers, some of our data, and descriptions of some of our programs.

2 Shape Description from Intensity Images

One of our goals for the last few years has been to develop the ability to generate segmented, volumetric descriptions from a single intensity image. Use of a single intensity image is attractive due to the ease and simplicity with which such data can be acquired. However, use of a single image poses many difficulties:

1) We must separate an object from the background (and from other objects). In general, objects are not characterized by simple image properties, such as uniform intensity or color. Rather, object segmentation must function in the presence of real image phenomena such as noise, surface markings, shadows and occlusion.
2) We need to recover 3-D structure from the 2-D image.
3) We need to segment an object into parts and describe the parts and the relations between them.

Our approach uses a shape description scheme based on generalized cylinders to model the component parts of a complex object and on generic joint relationships between the parts. It is based on an analysis of rigorous projective properties (geometric invariant and quasi-invariant, and structural properties) of large subclasses of generalized cylinders and of their joint relationships. It detects potential visible object fragments, groups them and verifies them, by searching for evidence of the expected projective properties among image boundaries. The obtained descriptions are then used to recover 3-D shape of each component of the part/whole hierarchy.

USC IMAGE UNDERSTANDING RESEARCH: 1993-1994

R. Nevatia, K. Price and G. Medioni*
Institute for Robotics and Intelligent Systems
Computer Science Department
University of Southern California
Los Angeles, California 90089-0273

Abstract

This paper summarizes the USC Image Understanding research projects and provides references to more detailed sources of information.We have undertaken a broad range of research with major efforts directed toward developing image understanding techniques to infer three-dimensional shape descriptions from range data and from intensity images, building detection and change analysis supporting the RADIUS program, incorporation of knowledge representation techniques in computer vision programs, and an effort in implementing vision algorithms on parallel systems.

1 Introduction

This paper summarizes the progress in our research projects since the last Image Understanding Workshop. Much of this work is described in detail in other papers in these proceedings, with this overview giving only a brief description of the detailed efforts.

Our research covers a broad range of separate tasks in image understanding, but the different tasks are highly inter-related and share many common techniques. The major task areas are three-dimensional descriptions from single images, 3-D descriptions from range data, navigation using trinocular stereo, building detection and change analysis in the RADIUS program, knowledge representation for computer vision, and parallel implementations of vision algorithms.

We first give a very brief outline of our work over the last 18 months, then we present results from many of the individual projects with references to the papers that appear elsewhere in these proceedings.

* This research was supported in part by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F49620-90-C-0078, Grant No F49620-93-1-0620, and F49620-94-1-0431; the Rome Laboratories under Contract No. F30602-93-C-0064; or the U.S. Army Topographic Engineering Center under Contract No. DACA76-93-C-0014. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

Object Descriptions from Intensity Images

This is one of the most difficult, but important, tasks in IU. Scene segmentation is difficult, as different types of features such as object boundaries, surface orientation isocontinuities, surface markings, shadows and noise cannot be directly distinguished. We use a process of perceptual organization to compute the higher-level descriptions in such cases. To be generic, perceptual grouping methods must use general methods. One method propagates the influence of local features over large vector fields and finds the most salient features. For higher level groupings, we use methods based on utilizing projective properties of contours of a class of objects. We have chosen generalized cylinders (GCs) as suitable volumetric representations. A few types of GCs (and their combinations) can represent a large fraction of the man-made objects in our environment. In recent research, we have developed some very powerful invariant (and quasi-invariant) symmetry properties of projected contours of GCs. Our studies indicate that we can use these properties to segment objects, fill gaps even in presence of occlusion and infer 3-D shapes from monocular images. Section 2 describes this work.

Analysis of Range Images

The goals of our effort in Range image understanding are to generate rich descriptions from sensed 3-D data. These descriptions should be segmented and capture both the volumetric and surface information related to objects. One of the applications is the automatic generation of 3-D models from multiple range images. We describe below in more detail four specific aspects of our research where we are studying:

- the integration of multiple range images into a surface representation of the object (in Section 3.1)
- a framework to handle complex objects with holes or multiple components (in Section 3.2)
- an approach to generate volumetric part descriptions from a surface representation of an object (in Section 3.3)